

Description

Voice recognizer and operating method therefor

- 5 The invention relates to a voice recognizer according to the preamble of claim 1 and an operating method therefor.

Having long secured itself a permanent and constantly growing application area in the input of text to office applications
10 running on PCs, voice recognition is also making increasing inroads in the control of technical devices. Both in ultra-miniaturized and at the same time computerized hand-held electronic devices, in particular mobile phones and PDAs, and in technical devices that are meant to involve minimum
15 attention and concentration from the user to operate, such as the various technical devices in a moving car, this type of voice recognition together with voice control based thereon can find useful potential applications. In the former type of devices, the area available for control actions has actually
20 become so small that the numerous possible functions can only be implemented very inconveniently using traditional keypad or touch-screen entries (and almost not at all for people with poor sight). In areas of use in which the attention of the user must remain focused on other things (e.g. road traffic),
25 the introduction of voice control not only increases convenience but greatly improves safety.

In voice recognition, a lexicon containing the words to be recognized is required. In the case of phoneme-based voice
30 recognition, these words are transferred by means of a text-to-phoneme technique into a phonetic transcription and saved in the vocabulary. During the recognition process, a search for the best path through the phoneme strings contained in the vocabulary is made using the Viterbi algorithm as it is known.

Details of the established voice recognition algorithms are given in the relevant technical literature.

Highly computerized technical devices of the aforementioned type (PDAs, hand-held PCs, mobile phones, vehicle audio systems, on-board computers etc.) have user-interfaces or MMI structures that are derived from PC user interfaces. There are a large number of applications installed that need to be controlled in a suitable way, and also in more complex devices in a specific sub-level of a logical hierarchy. In traditional devices of this type, menu-based control is provided for this purpose that can be executed by the user using soft-key entries.

When selecting an application by voice input, the program names of the available applications are contained in the lexicon. Once a name is recognized, the relevant program is executed or the application started. To do this, the program name and the program path must be saved in a suitable format.

According to the state of the art, the individual program names are hard-wired to the corresponding recognition results (the words in the lexicon). This can be specified in an additional file, or permanently defined in the source code of the program. Both methods have essential disadvantages, which are described below:

- When working with an additional file there is the problem that it can be seen by the user and consequently can also be modified. Even binary formats or write-protected files offer no effective protection against changes. This can lead to discrepancies between the vocabulary used and the word list or program list, with the consequence that the application may respond incorrectly.

- When the voice expressions acting as control commands are defined in the source code, it is not easy to make further changes to the vocabulary. The source code would need to be re-compiled and shipped every time changes in the program names occurred.

- The crucial disadvantage of the technique used up to now is the non-existent or inadequate system expandability. At present, it is not possible for the user to record his own commands or applications for inclusion in the automatic voice recognition, at least not without the risk of a fault in the originally programmed configuration of the voice recognizer.

The object of the invention is thus to provide an improved voice recognizer and method for its operation with which said device can be configured more flexibly in order to include the user's own control commands or applications.

This object is achieved in its device aspect by a voice recognizer having the features of claim 1, and in its method aspect by an operating method having the features of claim 6.

The invention includes the fundamental idea of providing a user interface constructed using links for the voice control of applications or for suitable handling of files. The organization principle of the links enables programs or files in different hierarchy levels to be opened easily in a structured way without a rigid assignment needing to be defined and programmed in advance.

The list of words to be recognized (the lexicon) is defined by the contents of a specific file directory which contains links (shortcuts) to the programs or files present. The name of the link specifies the word to be recognized, and the program or file to which this link points specifies the action to be

performed. In converting the name, one should note that only the partial string in front of the first dot is used as a command. The vocabulary is generated when the recognizer program is started. This allows a flexible response to changes
5 in the application structure or file structure. As soon as a word is recognized, the relevant link is actuated and the required action executed.

Advantages compared with the technique used up to now lie in
10 the flexibility regarding words and actions, and the simple creation and modification of a complex recognizer vocabulary. New commands can be added to the existing vocabulary in a simple and familiar way. A shortcut to the required program or file merely needs to be created in the file directory. Under
15 Windows, for example, a shortcut can be created easily via the context menu.

This illustrates a further advantage: since the file system takes over the management of commands and actions (name and
20 destination of the shortcut), no additional program is required for managing the vocabulary. If a command is meant to be deleted, the link is simply deleted.

Since modern operating systems allow links to files as well,
25 documents can also be opened by voice command.

In a preferred embodiment, the file directory includes a plurality of sub-directories in at least one subordinate hierarchy level, the directory names forming a first and if
30 applicable further, active partial vocabularies of the voice recognizer lower down the hierarchy.

By using sub-directories in the file directory, structured voice commands to open programs and files can be generated in
35 the simplest way. For instance, all links to pieces of music

are saved in a sub-directory "music". The word "music" is held in the active vocabulary in the first stage of recognition. If it is recognized, the vocabulary is switched (e.g. by language model), and the links contained in the "music" sub-directory are now held in the active vocabulary.

In particular, each program or file is assigned from a sub-directory a voice command composed of multiple connected parts that contains the names of the links from the file directory and each subordinate sub-directory leading to the program or file.

Complex voice commands can be created and edited in the simplest way using this method. Existing directories containing shortcuts, such as the Windows start menu, can now be operated simply by voice control because all necessary information is already there.

This method is a further development of shortcuts to programs (for example Windows PC) and the hard-wired voice recognizer resources. In this method the recognizer resource is provided automatically by the creation of a link, i.e. the name of the link can be processed by the recognizer immediately afterwards.

In general, any files and programs can be opened by voice command once they have been copied into the special directory. It also makes no difference whether a music title, c++ file, text document or program is involved. By saving a link in the special directory, the file is opened by the default program configured. For example, a document with the .doc extension is opened automatically by the Word program (as when double clicking on the file in traditional PC entry).

The aspects of the invention explained above appear both as the device aspect of a voice recognizer and as aspects of the operating method thereof, particularly since it is typically implemented in a suitable mix of hardware and software components.

Two ways of recording a word in the recognizer lexicon are given below:

(1) Recording by a program call via the context menu for the required application. In this case the context menu contains two program calls (Add and Remove). Add adds the relevant program/file and Remove displays the list of programs/files that can currently be selected by voice selection.

(2) Using "drag'n'drop" to copy the link to the required application into the special folder. In this case, in order to remove a program, one must switch to the relevant directory and delete the required link from the directory by "deleting".

The implementation of the invention is not limited to the examples and aspects described above, but is possible in numerous variations falling within the bounds of proper action.